ISDA
Safe,
Efficient
Markets

May 2026

Capital Models Benchmarking: A Framework for Counterparty Credit Risk Internal Models

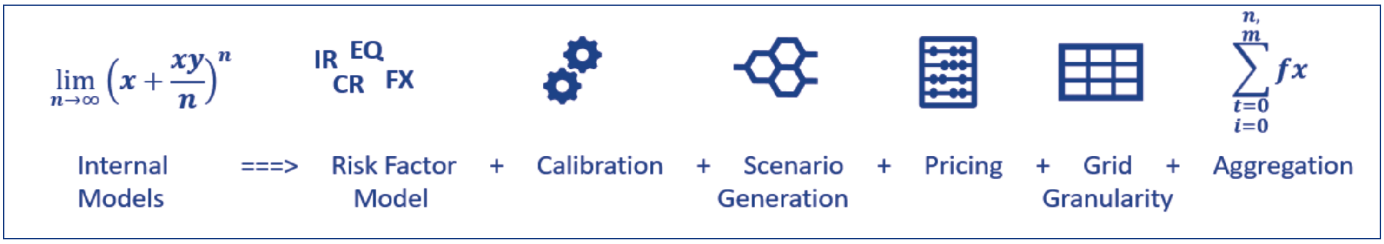
CONTENTS

INTRODUCTION

When firms implement capital models in line with supervisory standards, a range of interpretative and implementation choices inevitably arise. These choices reflect differences in modeling approaches, data availability, system architecture and risk management practices, and can lead to variation in model design and outcomes across the industry. The degree of flexibility may be relatively limited or more substantial, depending on the specific capital framework – such as credit risk, market risk or counterparty credit risk – and on whether the model is based on a standardized or advanced approach.

Internal models for counterparty credit risk (CCR) are an advanced regulatory approach for measuring CCR exposures. The resulting exposure measures are used for both risk management purposes and for the calculation of exposure at default (EAD) in default risk capital. These models are typically based on Monte Carlo simulation techniques, requiring firms to make a series of choices relating to inputs, models and techniques when developing and maintaining their internal models.

Figure 1: Implementation features of an internal model that firms must select from



All these modeling choices have varying levels of complexity with respect to approach, data requirements and associated costs and benefits. Firms typically assess the soundness of these choices through internal model validation practices – for example, by referencing previous model versions, comparing outputs to front-office pricing models or analyzing sensitivities.

Benchmarking to peers complements these internal assessments by extending model validation to external comparison across firms' internal models using hypothetical portfolios. In this paper, benchmarking refers to structured, like-for-like comparison of results and modeling choices across institutions, providing an external reference point that cannot be obtained through firm-specific analysis alone.

A robust peer benchmarking framework provides benefits to both firms and supervisors by introducing transparency and comparability of internal model outcomes across institutions. By answering cross-sectional questions, benchmarking supports supervisory dialogue, promotes convergence towards supervisory expectations and strengthens confidence in the credibility and consistency of internal models.

This paper puts forward a benchmarking approach that can be applied to the internal model for CCR. This is an industry approach developed with engagement from the UK Prudential Regulation Authority (PRA) as part of a wider regulatory initiative[1]. The paper sets out what benchmarking is, the questions it answers and how this complements model validation practices. It then puts forward the criteria for a robust capital models benchmarking framework, using

---

[1] PRA statement on supervisory benchmarking exercise relating to capital internal models, Bank of England, May 18, 2021, www.bankofengland.co.uk/prudential-regulation/publication/2021/may/pra-statement-supervisory-benchmarking-exercise-relating-to-capital-internal-models

ISDA's experience in the development of its award-winning standardized approach benchmarking solution[2], which includes ISDA Analytics. Finally, it illustrates how this approach has been leveraged to benchmark internal models for CCR[3].

---

[2] ISDA Wins Risk's Innovation in Technology Award For ISDA SA Benchmarking, ISDA February 17, 2022, www.isda.org/2022/02/17/isda-wins-risks-innovation-in-technology-award-for-isda-sa-benchmarking/

[3] This section describes the benchmarking approach developed in the most recent pilot of the ISDA UK Prudential Regulation Authority's internal model method benchmarking initiative. All results and examples are illustrative and do not include specific firms' submissions. Furthermore, the approach is evolving as the initiative progresses and therefore subject to revision and improvement

BENCHMARKING AND MODEL VALIDATION

Benchmarking is a model validation technique that extends traditional internal validation by introducing structured comparison across firms. In this paper, benchmarking refers to peer benchmarking: the quantitative or objectively qualitative comparison of a firm's model outcomes and modeling choices across institutions under common specifications using hypothetical portfolios, with the objective of providing an external reference point for assessing model plausibility and consistency.

In regulatory capital contexts, benchmarking is often associated with a hypothetical portfolio exercise (HPE), whereby firms calculate exposures or capital requirements for a prescribed set of stylized trades and portfolios, enabling like-for-like comparison of outcomes. While an HPE is a central and well-established component of peer benchmarking, a robust benchmarking approach is not limited to comparison of final capital or exposure measures alone. Meaningful benchmarking can extend to intermediate results, distributions, derived metrics and ratios, as well as qualitative aspects of model design and implementation, such as calibration methodologies, modeling conventions and parameter choices. Considering this broader set of information is essential for interpreting observed differences in outcomes and distinguishing between legitimate modeling choices and weaker practices.

Traditional model validation practices are inherently firm-specific. They answer questions related to the coherence (documentation of assumptions, limitations, risks), correct implementation (compared to requirements/specifications), and compliance with regulation (impact analysis of results, control checks on inputs/outputs, etc) of the model within the firm. To support this, firms use independent, alternative models that are developed or maintained alongside the firm's primary model for the purpose of assessing the reasonableness, robustness and limitations of the primary model's outputs. These firm-specific validation practices also include backtesting as a core component of the internal model method (IMM) validation framework, providing an empirical assessment of model performance by comparing model outputs against realized market outcomes over time.

While these internal validation tools are essential, they remain self-referential by construction. They cannot address cross-sectional questions such as whether a firm's model outcomes are extreme relative to peers facing the same risks, or whether specific modeling choices are aligned with prevailing industry practices. Benchmarking addresses this gap by introducing cross-firm comparability and enabling analysis of dispersion across institutions.

In this way, benchmarking complements internal model validation by extending the assessment from internal consistency and compliance to external coherence and industry plausibility. By answering cross-sectional questions, benchmarking strengthens confidence in internal models and enhances the credibility of model outcomes for both firms and supervisors.

A ROBUST BENCHMARKING FRAMEWORK

A foundational design principle for a robust benchmarking framework is that it must have clear objectives upfront to avoid the exercise being simply a league table of outcomes. It requires a focus on the comparability of outcomes across participants, ensuring that assessments are not confined to a single summary metric but instead consider a broader set of outputs, sensitivities and distributional characteristics. By examining results across multiple dimensions, the framework can distinguish structural differences in modeling approaches from superficial variations in headline figures.

Transparency and interpretability are equally critical. Results must be presented in a manner that is clear, accessible and supported by intuitive visualizations that explain the drivers of variation. A well-constructed framework incorporates explicit materiality criteria to separate statistical noise from economically meaningful deviations, thereby directing attention to areas of genuine concern.

Robust validation and verification processes are essential to ensure that comparisons are performed on a true 'apples-to-apples' basis, with consistent data definitions, scenario specifications and methodological assumptions. Furthermore, the design should be scalable and future-proofed, capable of accommodating new risk factors, instruments, portfolios and increasing dimensionality – such as additional tenors or simulation paths – without significant re-engineering. Automation of infrastructure and processes is key to supporting growing participation efficiently while containing costs.

Finally, benchmarking must be actionable, with conclusions that provide practical insights to inform both firms and supervisors.

ISDA FRTB-SA Benchmarking

ISDA has coordinated industry benchmarking exercises for regulatory standardized capital models over several years, most notably through its Fundamental Review of the Trading Book Standardized Approach (FRTB-SA) benchmarking initiative, although FRTB-credit valuation adjustment and the standardized approach for counterparty credit risk (SA-CCR) have also been benchmarked as part of this initiative.

Launched in 2018 with 15 UK-supervised banks and support from the UK PRA, the FRTB-SA initiative has expanded significantly in scale and geographic reach to cover 81 participating banks, including 24 global systemically important institutions. It has been used to support regulators around the globe as part of their own benchmarking exercises, and used by members to support data quality improvements and inform their submissions to the European Banking Authority's annual benchmarking exercise for market risk internal models[4].

The benchmarking process comprises two components: an aggregation test and an HPE. The aggregation test checks the accuracy of the capital model implementation by standardizing the inputs to the model, thereby ensuring any results are directly comparable by construction to what is expected. In the HPE, banks conduct end-to-end capital calculations with a set of hypothetical trades grouped into specific portfolio combinations. This component allows for a comparison of outcomes and an understanding of the variability across firms with the same risk definition.

To support each process, ISDA has developed in-house tools that participants use to inform the accuracy and consistency of their implementation of the FRTB-SA. The ISDA Capital Common

---

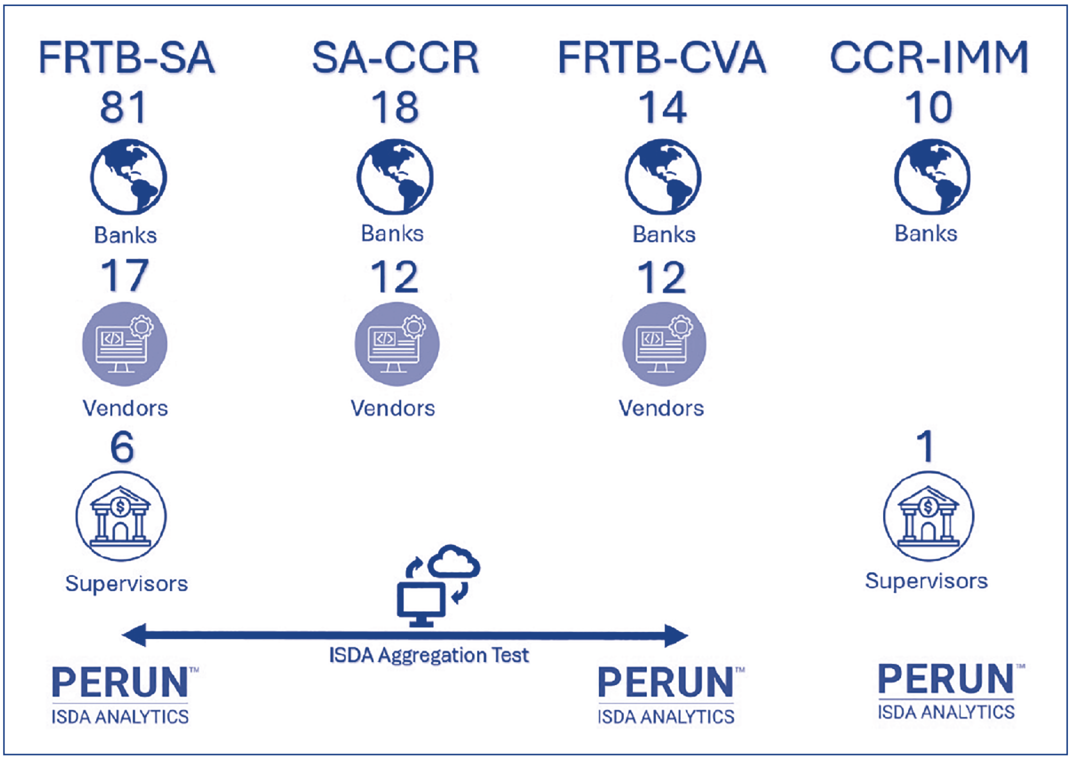
[4] ITS package for benchmarking exercises, European Banking Authority, www.eba.europa.eu/activities/single-rulebook/regulatory-activities/supervisory-benchmarking-exercises/its-package-benchmarking-exercises

Risk Interchange Format (CRIF) data standard has been leveraged to create a single data standard for risk data used in capital model calculations across participating firms. The HPE is supported by ISDA Analytics, an online platform for secure data loading, validation, results analysis and on-demand visualization from users.

Crucially, via capital variability analysis (CAVAR), the analytics can explain the drivers of dispersion. This allows banks and regulators to identify and understand specific causes of variability, down to individual portfolios and risk types.

ISDA has sought to devise an extensible framework that can be used to benchmark internal models for CCR. This has been done by building on the experience gained from the development and successful industry adoption of the standardized approach benchmarking solutions, with input from industry members and insight from the UK PRA. Such a framework would seek to achieve the objectives of a robust benchmarking framework by exploiting mutualized tools, data standards and industry engagement. The following section describes such an approach.

Figure 2: ISDA capital models benchmarking ecosystem



BENCHMARKING CCR INTERNAL MODELS: AN INDUSTRY APPROACH

The overarching objective of the benchmarking approach is to identify and understand sources of variability in internal model outputs across institutions. This is achieved by using standards, processes and tools to ensure a scalable, well-understood framework that delivers value for supervisors and participating firms.

The approach can be easily integrated into standard model supervisory processes rather than being a standalone activity. This is an important reason for firms to integrate benchmarking into business-as-usual systems and processes, rather than trying to mobilize resources each time.

The following sections of the paper cover the scope, activities involved, data requirements, benchmarking model metrics, data standards and supporting tools for benchmarking CCR internal models. Sample results and the associated analysis that helps explain the variability in the model metrics are also described to illustrate the usefulness of the approach for participants[5].

CCR Scenario Scope

The benchmarking scope covers both current end-of-day calibration and stressed window calibration scenarios of CCR exposures.

Benchmarking Activities

The benchmarking approach consists of three activities, all of which complement each other.

1. Industry Working Group

A regular industry working group of participating firms coordinated by ISDA supports the benchmarking activities and processes. The group's responsibility is to provide input into the design of a robust benchmarking framework for CCR internal models, execute the framework and analyze and report results that satisfy the objectives of benchmarking. As part of the governance, ISDA facilitates the discussions across the group to ensure a careful balance between transparency and data protection. The network effects of the working group provide significant benefits in understanding the feasibility of the approach, as well as what data requirements and features help to explain the variability seen in the benchmarking of model metrics.

2. Qualitative Survey

Qualitative surveys are carried out across participating firms to understand model implementation features and assumptions. The questions and scope of possible answers are devised by the working group. Examples of topics surveyed include risk factor model parameters, calibration methods, collateral modeling assumptions and time horizon grid granularity. These surveys are useful to support the interpretation of quantitative results via qualitative categorization of data clusters, as well as benchmark industry norms with respect to risk factor models assumptions or implementation.

---

[5] All results and examples are illustrative only and do not include specific firms' submissions. Furthermore, the approach is evolving as the initiative progresses and is therefore subject to revision and improvement

3. Quantitative Benchmarking (including HPE)

Benchmarking CCR model metrics across participating firms is the core activity of the framework. To facilitate this quantitative benchmarking, an HPE approach is a key element. Defining hypothetical portfolios with uniform legal agreements, collateral properties and trade elements across participating firms ensures a degree of control and standardization to generate a like-for-like comparison of CCR model metrics (final and intermediate) as much as possible without enforcing artificial uniformity.

Using hypothetical portfolios allows for the benchmarking of well-understood counterparty exposure metrics, such as effective expected positive exposure (EEPE) and expected positive exposure (EPE), that do not require transformation from submitted data and material variability in results is relevant and consequential for CCR default capital. The HPE approach is also extensible in that it supports a building block approach where the scope of risk factors, payoffs and portfolio features starts simple and then incrementally the complexity is increased as required by participating firms or supervisors without significant changes to the approach of using hypothetical trade or portfolio settings.

The following table summarizes the scope of the HPE in the most recent CCR internal models benchmarking exercise, in which 48 portfolios were defined based on these properties.

Trade Type	Risk Factor	Maturity	Legal Agreements	Collateral Type
Interest rate swap	EUR, USD, GBP	1y, 5y, 10y	ISDA Master Netting Agreement (MNA) ISDA MNA + Credit Support Annex (CSA)	USD cash
FX forward	USD/JPY	1y, 2y, 5y	ISDA MNA ISDA MNA + CSA	USD cash
Zero coupon inflation swap	EUR, USD, GBP	1y, 5y, 10y	ISDA MNA ISDA MNA + CSA	USD cash

Data Requirements

The benchmarking exercise relies on a rich data set, comprising both end results and intermediate model outputs, providing multiple perspectives on firms' CCR internal model behavior. This breadth of information enables deeper analysis to interpret any variability observed in the benchmarking metrics shared with participants, whether arising from exposure aggregation, valuation dynamics, collateral modeling or risk factor simulation.

The quantitative data set comprises:

- CCR exposure measures: EEPE, expected exposure (EE), SA-CCR exposure at default (EAD);
- Trade / instrument measures: mark-to-market, cashflows, notional(s);
- Monte Carlo simulation distributions for:
  - Portfolio values per simulation path at each grid point;
  - Collateral values per simulation path at each grid point;
- Risk factor distributions per simulation path at each of the firms' grid points.

Data Validation and Integrity Checks

To ensure that conclusions drawn from analysis of benchmarking results are reliable and well-founded, it is essential that the various data elements are internally consistent with one another. As a result, a series of validation checks are applied to confirm alignment between submitted measures and to establish confidence in the integrity of the data used for benchmarking and explanatory analysis.

The core validation checks are:

- Recalculation of EEPE from firms' submitted EE profiles to verify consistency between aggregated exposure measures and the underlying exposure term structure;
- Calculation of EE directly from firms' portfolio value distribution submissions to confirm alignment between reported exposure profiles and the simulated portfolio value dynamics produced by the internal models.

Data Standards and Tools

This section sets out the data standards and automated solutions used to support the industry approach and the objective of a robust, industrialized process that is repeatable and scalable.

Capital CRIF Data Standard

There is a strong need for a standard and repeatable way of describing CCR metrics on portfolios and trades. A key enabler of efficiencies in benchmarking processes and scalability in ISDA's approach is the ISDA CRIF, a standardized template for exchanging risk data. The IMM CRIF builds on ISDA's earlier CRIF standards, such as those developed for the ISDA Standard Initial Margin Model (ISDA SIMM) and SA-CCR. In particular, the IMM CRIF is an evolution of the SA-CCR CRIF, adapted to address the specific challenges of the IMM.

The capital CRIF standard has a simple, robust format, which can be read by automated processes as well as being capable of manual inspection. Whereas factor-based frameworks only require limited data points, IMM benchmarking requires the exchange of rich, simulation-based outputs.

The IMM CRIF has been designed to:

- Capture IMM-specific risk types, including netting sets and collateral agreements;

- Represent Monte Carlo simulation outputs in a structured, machine-readable way;

- Ensure that data from different institutions can be compared consistently and efficiently.

This common format allows large volumes of benchmarking data to be submitted, validated and aggregated in a scalable way, forming the backbone of a repeatable and industrialized benchmarking process. Full details of the format and data standard can be found in the latest version of the Risk Data Standards document, which is shared with all participants.

LSEG Open-source Risk Engine[6]

Benchmarking standardized models using CAVAR, supported by an in-built calculator of the FRTB-SA capital model, proved invaluable in understanding the drivers of dispersion and their

---

[6] Open Source Risk Engine, London Stock Exchange Group (LSEG), www.lseg.com/content/dam/post-trade/en_us/documents/post-trade-solutions/fact-sheets/open-source-risk-engine-factsheet.pdf? Open Source Risk Engine, LSEG, www.lseg.com/en/post-trade/solutions/advise/open-source-risk-engine?

material impact. The benefits of such an approach cannot be fully replicated for internal models benchmarking, given the increased complexity and degrees of freedom of model choices (Figure 1), which preclude a single implementation. However, having access to an independent CCR risk engine that computes equivalent exposure measures provides similar and significant benefits to analyze benchmarking results and support explanatory analysis.

On this basis, the London Stock Exchange Group (LSEG) Open-source Risk Engine (ORE) has been integrated into the benchmarking framework for CCR internal models. ORE is a transparent, open-source software framework for pricing, XVA and risk analytics developed by market practitioners and academics and maintained under the LSEG umbrella . ORE has been adopted by many leading financial institutions across the globe as part of their pricing and risk infrastructure, including business critical processes such as ISDA SIMM calculation and reconciliation. Its broad analytical coverage across asset classes, combined with an extensible and openly documented codebase, makes it well suited as an independent point of reference alongside firms' proprietary internal models.

Within the benchmarking exercise, ORE provides a consistent analytical baseline that supports validation of trade definitions, initial market valuations and scenario-based analysis. While it is not intended to replicate firms' internal models, its use enhances the interpretability and robustness of benchmarking results by providing a transparent and common reference for comparison and explanatory analysis.

ORE is leveraged in several aspects of the HPE component of the benchmarking framework:

Trade Definition: Utilizing ORE trade definition schemas across the relevant trade types, it is possible to provide a more complete definition of the trades to be booked by participating firms upfront. This is extremely useful for the more complex trade types. Moreover, by using the valuation functionality in ORE, it was possible to prescribe market variables such as fixed rates and strikes for trade types upfront for all firms.

Trade Booking Validation: The valuation functionality in ORE provides a reference valuation for each trade in the HPE that can be compared to firms' submissions as part of the initial market valuation check, which is performed to ensure consistency of trade booking across participating firms. Where submissions from firms are materially different, ORE-supporting reference data and market data can be shared with firms to understand the root cause of the difference and whether re-booking/re-valuation is required to align consistency, or whether differences are to be expected due to known differences in inputs.

What-if Analysis: ORE supports targeted what-if analyses in which specific modeling or configuration choices are varied in a controlled manner to assess their expected order-of-magnitude impact on exposure measures. These analyses provided useful reference points for interpreting observed dispersion and understanding the sensitivity of benchmarking results to specific modeling assumptions. Three examples of the what-if analysis performed, and conclusions drawn, are illustrated below.

1. Base Currency Configuration: One of the what-if analyses performed was to understand if the simulation base currency differences across participating firms would result in material differences in output measures of the risk factors, trade types and payoffs in scope.

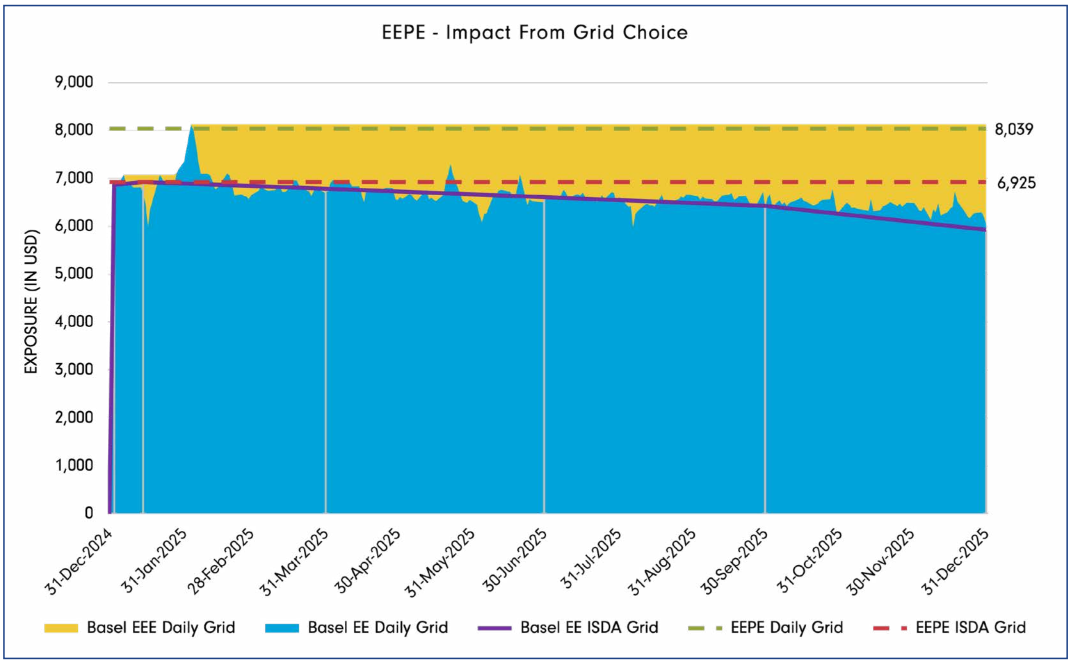
Figure 3: Comparison of EEPE as a percentage of notional ($/€/£1 million), calculated with EUR/USD base simulation currency

For the non-FX-related trade types, the results clearly indicate there is no material impact on EEPE because of base simulation currency differences across any of the portfolios, irrespective of the maturity of the trade. It is therefore reasonable to conclude that any material variability across participating firms with base currency differences could not automatically be attributed to such an explanation for these portfolios. For the FX portfolios, the maximum EEPE impact is 13%, so some impact from base currency differences could be relevant when interpreting results.

2. Grid Granularity Configuration: Grid granularity varies across firms, both in terms of the number of grid points used and the mapping of generic tenors (two-week, one-month, three-month, etc.) to specific calendar dates. These differences can materially affect exposure profiles, particularly when cashflows or resets occur between grid points. To assess the extent to which grid granularity alone can explain variability in exposure measures, targeted what-if analyses are performed using alternative grid configurations for the same portfolios. These include grids that reflect common industry implementations, as well as finer-grained reference grids. Exposure measures are then recalculated and compared to a baseline configuration. This analysis provides an indication of whether observed differences in firms' EE and EEPE profiles are plausibly attributable to grid granularity – for example, where a sparser grid fails to capture the impact of a material cashflow occurring between standard grid points.

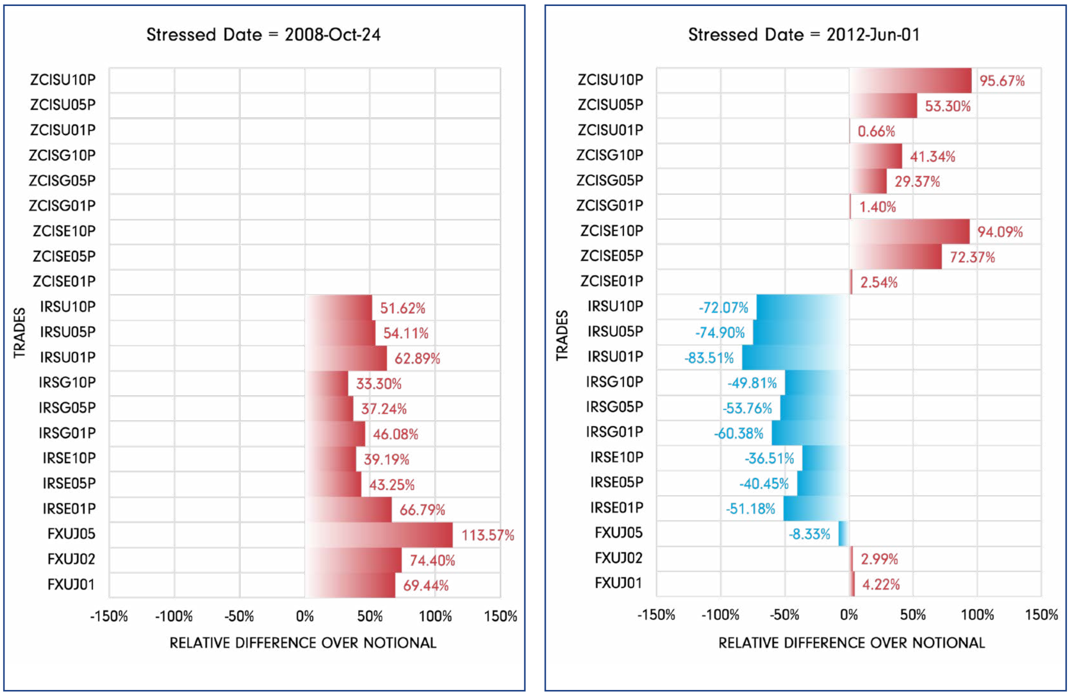
Figure 4 illustrates a simple example in which a daily grid captures a cashflow-driven spike in exposure that is partially or fully missed under coarser monthly (after two weeks) or quarterly (after two weeks) grids.

Figure 4: EEPE variability on a 10-year US$ interest rate swap because of grid granularity differences



3. CCR Stress Calibration Period Definition: Another area where differences across participating firms exist is the determination of the stressed expected positive exposure (SEPE) calibration period. Evaluating the stressed calibration results across the cohort is not an exact like-for-like comparison, given these differences. It is operationally very difficult for firms to align to a single stress period definition. Using ORE to produce a set of results across different stress periods within the same risk engine helps to determine the variability coming from the calibration period choice seen across firms' submissions, which can be used to inform the analysis of the results. Figure 5 shows a simplistic example of the EEPE impact on the HPE portfolios against the current calibration for two stress calibration dates: October 24, 2008 and June 1, 2012 that was run on ORE. The first stress period occurred during the global financial crisis and the second is associated with the eurozone crisis. These two periods are commonly used across participating firms.

Figure 5: Comparison of EEPE results by asset class and maturity for all portfolios compared to base calibration of 2008 or 2012 stress calibration choice



Inflation trades were not run for stressed calibration date October 2008 due to market data constraints

These results show a significant impact for rates and FX for EEPE, depending on the choice of stress period[8]. These results could be used to understand whether the relative magnitude between base and stressed EEPE of firms' results aligns to what is seen above. For example, for the five-year FX forward, the 2008 stress calibration firms have a stressed EEPE (SEEPE) significantly greater than base, whilst the 2012 stress calibration firms have a similar sized or slightly smaller EEPE compared to base. If such results are mirrored, the stress calibration window could be a significant indicator of the variability in stress SEEPE.

To ensure the results of any analysis using ORE were reliable, ORE data points were benchmarked against the firms' submissions for the portfolios, risk factors and payoffs. Exposure profiles and other relevant metrics were in line with the cohort of other benchmarking firms.

Benchmarking Results

The benchmarking outcomes, together with illustrative themes and findings from applying the framework in the most recent CCR internal models pilot benchmarking exercise, are set out below.

Measures

EEPE is the primary metric for benchmarking CCR internal models because:

1. It provides a single, well-defined measure of exposure that enables direct, like-for-like comparison across firms, without the need for additional normalization or transformation;

---

[8] This is an example to illustrate the use case. For this to be comparable to firm-specific results, the calibration of stressed volatilities would have to align more closely to the method the firm used, which may be a different date in the window for implied calibration or, in the case of historical calibration, not a single date but an average across the window

2. It forms the basis of default risk capital calculations under the IMM framework, meaning that material differences in EEPE translate directly into material differences in capital outcomes.

EEPE was submitted by firms on their internal production grid points. Setting up a standardized grid was extremely difficult because not all firms were able to do so without significant efforts and outside specifying dates firms would interpret generic labels such as 'two weeks' or 'one month' differently against a production grid, nullifying the attempts at standardization. More importantly, any results from such a standard grid could be very different compared to production (see Figure 4 for an illustration of EEPE impact because of grid point differences).

Benchmarking the EEPE provides a useful starting point for comparing CCR internal models, but it does not capture all relevant exposure dynamics. Effectivization of the EPE profile can play a significant role in the final EEPE result so that variability of the EEPE can be explained with simple reasons, such as valuation differences across firms.

In addition, EEPE is limited to a one-year default horizon, which limits the visibility of longer-dated exposure behavior. As a result, differences in assumptions that influence the volatility and shape of simulated exposure distributions may be partially obscured in EEPE-based comparisons, even when these differences are materially relevant.

To provide greater insight into exposure dynamics, the benchmarking framework includes a comparison of EPE[9] profiles across firms. This allows for a comparison of the shape and curvature of exposures over time, extending the analysis beyond the one-year default horizon. Differences in observed EPE profiles can provide indications of how risk factor model dynamics such as mean reversion, drift and volatility assumptions vary across firms.

Further transparency on the impact of these risk factor model dynamics is obtained by benchmarking metrics derived from firms' submitted portfolio value distributions. These include measures such as the mean and standard deviation of portfolio values across grid points, which provide insight into how the level and dispersion of simulated values evolve over time across the cohort.

A single standard measure of the potential future exposure to benchmark the tail of the distribution is also included in the benchmarking approach. Incorporating a tail-focused metric complements the benchmarking of average exposure measures by providing additional visibility into differences in the extreme outcomes of simulated exposure distributions.

In addition to benchmarking measures output directly from firms' models, the framework includes benchmarking of selected ratios to enhance interpretability and context. These include benchmarking the ratio of:

1. EEPE end-of-day calibration / EEPE stressed calibration;

2. EEPE internal model / SA-CCR normalized for alpha.

Benchmarking such ratios supports the assessment of relative conservatism and sensitivity to calibration choices across firms.

---

[9] EPE is used interchangeably with EE

Defining Material Differences

Outlier categorization is a structured analytical step within benchmarking that seeks to identify and classify model outcomes that deviate materially from peer results under common assumptions. Rather than serving as a judgment on model correctness, outlier categorization is used to prioritize investigation and focus analytical effort where it is most informative. By systematically grouping results into peer-consistent ranges and outlier categories, benchmarking moves beyond ranking to enable targeted explanation, constructive challenge and meaningful interpretation of dispersion across firms.

A practical starting point for defining materiality thresholds is to use simple, transparent, statistical rules, anchored to a central reference value such as the median of the cohort. In the pilot framework, two commonly used approaches were considered:

- Standard median fixed bands around the median (eg, flagging values above 150% or below 50% of the median);

- Median absolute deviation (MAD): thresholds that adapt to the dispersion of the data set (eg, median ± 2×MAD, where MAD = 1.4826 × median(|xi – median(x)|))

While straightforward, fixed-band thresholds can behave unpredictably when values are close to zero, and MAD-based thresholds are often more robust to heterogeneous dispersion. Accordingly, MAD-based thresholds were predominantly used in the IMM pilot.

Clustering is a key challenge for median-anchored thresholds. Both fixed-band and MAD-based approaches implicitly assume that the cohort can be represented by a single ‘central’ value around which variability is measured. In practice, benchmarking results can exhibit multiple clusters, driven by reasonable and relatively common modeling choices (for example, different but acceptable calibration methodologies or conventions). In such cases, applying a single, median-centred threshold may incorrectly classify an entire smaller cluster as ‘outliers’, even when those observations are internally consistent and reflect a coherent modeling approach that is present across several firms. This risk reinforces the importance of using qualitative information (eg, survey responses on modeling choices) to support meaningful grouping of results prior to applying mechanical thresholds, and to avoid over-interpreting dispersion that primarily reflects legitimate methodological heterogeneity.

In the current framework, mechanical thresholding is therefore combined with qualitative survey information to support clustering and interpretation. This allows thresholds to be applied within more homogeneous peer groupings where appropriate, improving the signal-to-noise ratio and reducing the likelihood that legitimate methodological choices are treated as anomalous. Determining robust approaches to clustering and outlier categorization remains an area of ongoing development as the initiative evolves.

Ongoing Research: Multi-measure and Machine-learning Approaches

In parallel to median-based thresholds, research is ongoing into the use of machine-learning techniques for outlier detection that use the joint set of measures available (eg, EEPE, EPE profile features, distribution-based metrics, tail measures and ratios), rather than relying primarily on a single headline metric. The potential advantage of such approaches is an improved ability to recognize multiple clusters and to identify observations that are inconsistent across several dimensions, even when they are not extreme on EEPE alone. The trade-offs are that the resulting outlier definition may be less transparent (because it depends on algorithmic models rather than simple closed-form thresholds) and may require recalibration for each exercise to reflect changes in portfolios, measures or the participating cohort. For these reasons, multi-measure techniques are best viewed as complementary to transparent thresholding – supporting triage and investigation – rather than as a replacement for explainable, governance-friendly materiality criteria.

Selected Benchmarking Results and Insights

This section presents selected results and insights produced by the benchmarking framework. The examples shown are illustrative and use stylized charts rather than firm submissions. They are not intended to be exhaustive, but to demonstrate the types of comparative insights the framework enables and how different sources of information can be combined to support interpretation of the results.

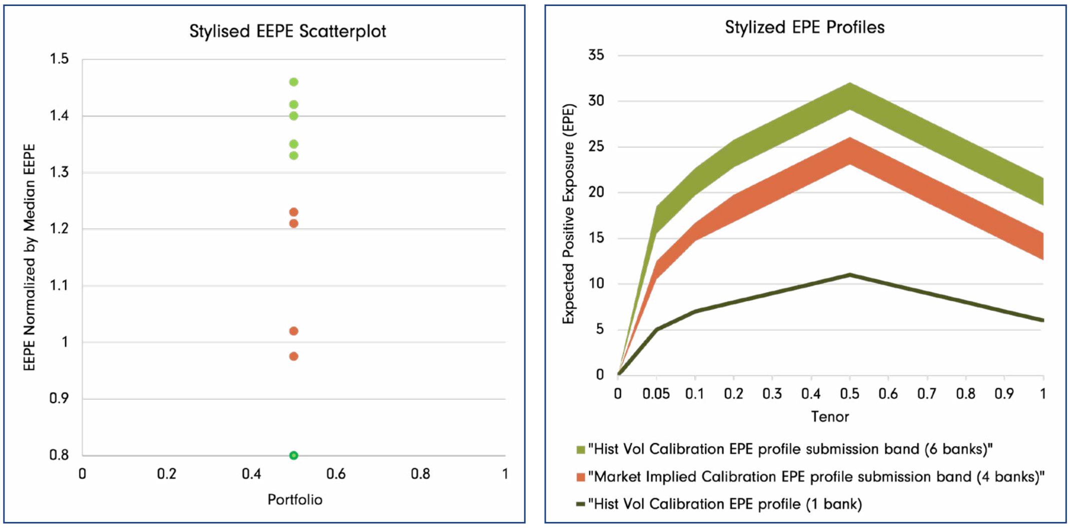
The emphasis is on explaining observed variability across firms, rather than on ranking outcomes. By combining quantitative benchmarking measures with supporting qualitative and intermediate data, the framework facilitates more informed analysis of model behavior and helps distinguish between differences arising from legitimate modeling choices and those warranting further investigation.

Use of Qualitative Survey Information to Support Clustering and Outlier Interpretation

Quantitative benchmarking results are interpreted alongside qualitative information collected through structured surveys on model design and implementation choices. This combined approach supports identification of coherent clusters of results associated with common methodological choices and improves the interpretation of dispersion across firms.

As illustrated in Figure 6, categorizing submissions based on volatility calibration methodology reveals distinct groupings of EEPE results. When such categorization is applied, it becomes clear that higher EEPE values clustered around 1.4–1.5 × the median reflects a consistent modeling approach shared by multiple firms, rather than anomalous behavior. In contrast, a single historically calibrated submission at 80% of the median stands out as the true outlier within its relevant peer group.

Figure 6: Stylized scatterplot to illustrate benchmarking results for EEPE and EE profile submissions with volatility calibration categorization from the qualitative surveys



Combining quantitative and qualitative data in the analysis also provides value when comparing results across different points in time. Analysis performed during the most recent benchmarking exercise, together with earlier feasibility work, shows that the relative impact of historical versus market-implied volatility calibration is not stable over time. For example, in 2024, the median EEPE for historically calibrated submissions was approximately 165% of the corresponding market-implied median, whereas in 2022 the relationship was reversed, with historically calibrated results closer to 80% of the market-implied median.

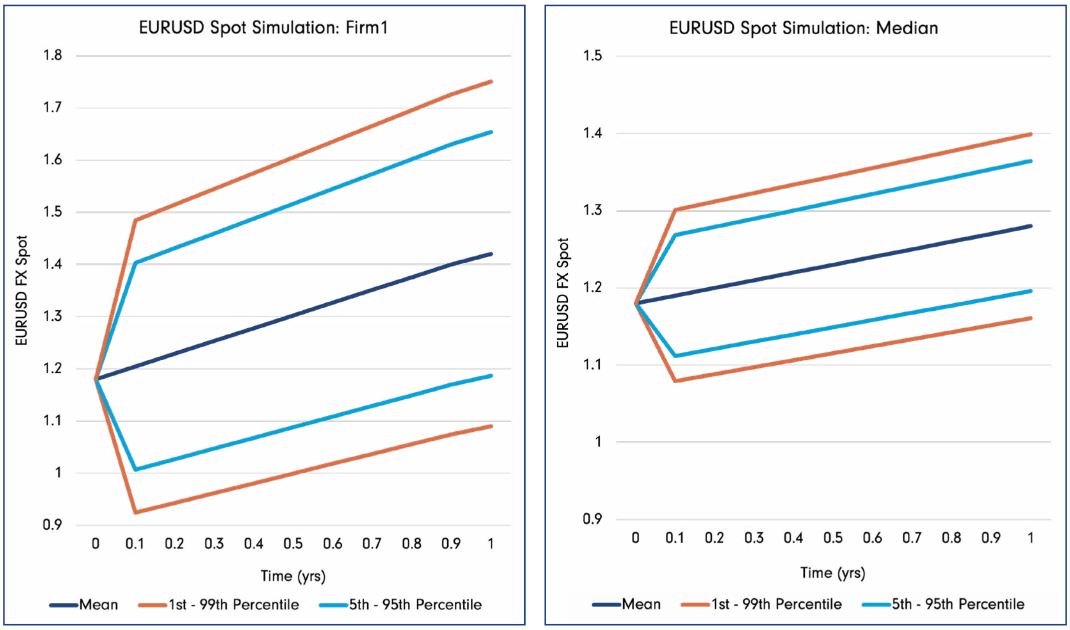
Presenting results in this way helps participants and supervisors to recognize that differences in benchmarking outcomes may reflect changing market conditions interacting with modeling choices, rather than persistent conservatism or weakness. It also illustrates the importance of avoiding static interpretations of relative outcomes and reinforces the need for context-aware analysis when drawing conclusions from benchmarking results.

Benchmarking Risk Factor Distribution

The benchmarking approach includes comparisons of simulated risk factor distributions submitted by firms. Analyzing risk factor distributions provides additional context for understanding how modeling choices – such as volatility calibration or distributional assumptions – translate into differences in downstream exposure measures. This helps participants assess not only the direction but also the order of magnitude of potential impacts arising from conservative or aggressive modeling choices, supporting a more informed and explanatory benchmarking discussion.

Figure 7 illustrates an example of risk factor distribution benchmarking for the EUR/USD spot rate over a one-year horizon. The chart shows the distribution submitted by a representative firm compared against a benchmark distribution constructed as the median of the corresponding risk factor distributions across participating firms. This median distribution serves as a reference point for assessing the relative dispersion and shape of individual firm simulations, rather than representing a separate or alternative modeling configuration.

Figure 7: Example of benchmarking results of a firm's EUR/USD FX spot distribution vs the median distribution



Presenting results in this way allows differences in distributional width, skew or tail behavior to be interpreted in relation to a peer-based benchmark. Wider percentile bands, for example, may indicate more conservative volatility assumptions, while tighter distributions may reflect lower calibrated volatility or stronger mean-reversion effects. Comparing firm-specific distributions to the median benchmark helps participants and supervisors to assess not only the direction but also the order of magnitude of differences implied by alternative modeling choices.

While the median peer distribution provides a natural and transparent benchmark within a cohort, alternative reference points may also be considered. Risk factor distributions generated by an independent third-party source, such as an open-source risk engine like ORE, could potentially be used as an additional benchmark. Such references can provide further context for interpretation, especially where the cohort itself exhibits multiple clusters or where an external point of comparison is useful to support explanatory analysis.

Used in conjunction with exposure and collateral benchmarking, risk factor distribution analysis strengthens the interpretability of results by linking differences in model outputs back to underlying assumptions about market dynamics.

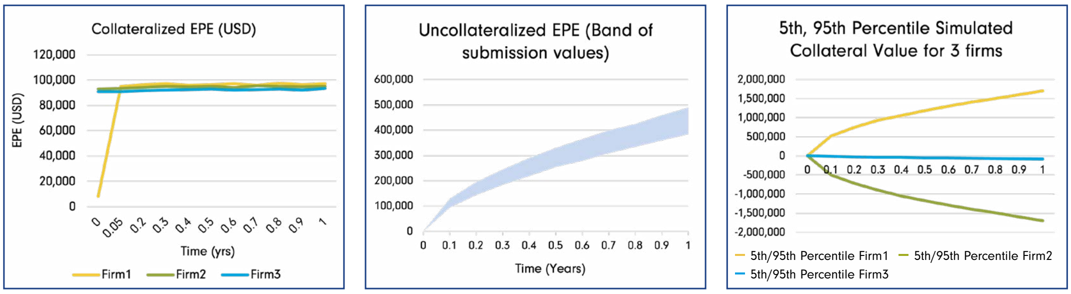
Using Collateral Distribution Results to Interpret EPE Profiles

Benchmarking of collateral distributions provides additional insight into differences observed in exposure profiles by making visible intermediate modeling outcomes that are otherwise difficult to infer from aggregate exposure measures alone. Analysis of collateral distributions can help distinguish whether variability in EPE profiles arises from differences in trade valuation dynamics, collateral modeling assumptions or timing conventions.

Figure 8 illustrates a stylized example based on a fully collateralized portfolio consisting of a single FX forward on USD/JPY. While the EPE profiles for three representative submissions are broadly similar over the time horizon, one submission exhibits a noticeably lower exposure in the initial period between T-0 and T+20 days.

Examining the corresponding uncollateralized EPE profiles shows limited dispersion, indicating that the observed difference does not originate from trade simulation or valuation.

Figure 8: Three stylized charts illustrating benchmarking results and explanatory benefits of collateral distribution data



Analyzing the submitted collateral distributions gives further insight into the driver of that difference. The average collateral value over time is broadly consistent across firms and the percentile bands indicate similar volatility and drift characteristics. However, one submission shows a clear delay in the evolution of the distribution of collateral values relative to the others, consistent with collateral modeled with a lag corresponding to the margin period of risk (backward-looking). This contrasts with submissions where collateral models are forward-looking.

By jointly analyzing exposure profiles, uncollateralized exposures and collateral distributions, the benchmarking framework can support clearer interpretation of observed differences by linking them back to specific modeling assumptions. More generally, the value of analyzing intermediate results lies in the ability to narrow the range of potential explanations and to inform more focused investigation. While attribution may not always be straightforward, particularly where multiple modeling choices interact, access to intermediate data can materially enhance understanding of benchmarking outcomes, rather than relying on aggregate exposure measures alone.

FUTURE DEVELOPMENTS

ISDA Analytics: CCR Internal Models Integration

ISDA Analytics provides the end-to-end workflow currently used to support benchmarking of standardized approaches, including secure data submission, validation, benchmarking execution and interactive analysis of results. This capability is to be extended to support CCR internal models benchmarking in a phased manner.

The initial phase will go live in the first half of 2026 and it focuses on data ingestion, validation and benchmarking of primary exposure measures, including EEPE and EPE profiles. Subsequent phases are expected to extend coverage to additional measures and explanatory analyses as the benchmarking framework continues to evolve.

LSEG ORE

The use of LSEG ORE as an independent analytical reference has demonstrated value in supporting interpretation of benchmarking results. Further extensions for its use, including additional what-if analyses, will continue to be explored where they can provide explanatory insight into observed variability. Specifically, ORE could be used to explore different market data configurations, as well as to benchmark risk factor distributions directly via LSEG ORE, thus isolating the distribution from other implementation choices such as grid granularity.

Future Benchmarking Phases

The industry benchmarking framework will continue to be refined in close collaboration with participating firms and supervisory authorities. Future phases will focus on extending analytical depth and scope, while preserving the building-block design that allows complexity to be introduced incrementally. This iterative approach aims to ensure that the benchmarking framework remains aligned with evolving modeling practices and regulatory expectations and continues to add value, both for firms' internal model governance and for supervisors' ongoing monitoring of CCR models.

CONTACTS

ISDA

Panayiotis Dionysopoulos
Global Head of Financial and Enterprise Risk
International Swaps and Derivatives Association, Inc. (ISDA)
25 Copthall Avenue, 3rd floor, London EC2R 7BP
Tel: +44 (0)20 3808 9700
pdionysopoulos@isda.org

Paola Rensi
Head of Capital Models Benchmarking
International Swaps and Derivatives Association, Inc. (ISDA)
25 Copthall Avenue, 3rd floor, London EC2R 7BP
Tel: +44 (0)20 3808 9700
prensi@isda.org

LSEG ORE

Xabier Anduaga
Partner, Quant Services
London Stock Exchange Group (LSEG)
10 Paternoster Square,
London, EC4M 7LS
Tel: +44 (0)7483 440100
xabier.anduaga@lseg.com

Alexis David
Senior Consultant, Quant Services
London Stock Exchange Group (LSEG)
10 Paternoster Square,
London, EC4M 7LS
Tel: +44 (0) 7423 370507
alexis.david@lseg.com

ABOUT ISDA

Since 1985, ISDA has worked to make the global derivatives markets safer and more efficient. Today, ISDA has over 1,000 member institutions from 79 countries. These members comprise a broad range of derivatives market participants, including corporations, investment managers, government and supranational entities, insurance companies, energy and commodities firms, and international and regional banks. In addition to market participants, members also include key components of the derivatives market infrastructure, such as exchanges, intermediaries, clearing houses and repositories, as well as law firms, accounting firms and other service providers. Information about ISDA and its activities is available on the Association's website: www.isda.org. Follow us on LinkedIn and YouTube.